

## **ON THE LIMITATIONS OF USING SITUATIONAL JUDGMENT TESTS TO MEASURE INTERPERSONAL SKILLS: THE MODERATING INFLUENCE OF EMPLOYEE ANGER**

JEREL E. SLAUGHTER  
University of Arizona

MICHAEL S. CHRISTIAN  
University of North Carolina at Chapel Hill

NATHAN P. PODSAKOFF  
University of Arizona

EVAN F. SINAR  
Development Dimensions International (DDI)

FILIP LIEVENS  
Ghent University

Many authors have suggested that situational judgment tests (SJTs) are useful tools for assessing applicants because SJT items can be written to assess a number of job-related knowledges, skills, abilities and other characteristics (KSAOs). However, SJTs may not be appropriate for measuring certain KSAOs for some applicants. We posit that using SJTs to measure *interpersonal skills* may lead to invalid inferences about applicants with higher levels of angry hostility (AH), and thus, AH should moderate the relation between interpersonally oriented SJTs and job performance. Three studies, using samples of health-care workers ( $n = 225$ ), police officers ( $n = 54$ ), and medical doctors ( $n = 92$ ), provided support for hypotheses in that that relations between SJT scores and performance criteria were significantly weaker among employees higher in AH compared to those lower in AH. In addition, none of the other facets of neuroticism tested (self-consciousness, anxiety, depression, immoderation, or vulnerability to stress) consistently moderated SJT validity, providing support for the uniqueness of AH. Implications for practice, and for future research studying the relations

---

We are grateful to Quinn McCarthy and Jennifer Anderson for their assistance with data collection. We also thank the I-O psychologists who assisted us by providing ratings of SJT items in Study 1. Portions of these data were presented at the 25th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, Georgia in 2010.

Correspondence and requests for reprints should be addressed to Jerel E. Slaughter, Department of Management and Organizations, Eller College of Management, University of Arizona, Tucson, AZ 85721; jslaught@eller.arizona.edu.

between interpersonal skills as measured by SJTs and job performance, are discussed.

Situational judgment tests (SJTs) present job applicants with situations they may encounter at work and ask them to make a choice among a number of possible ways to respond (Weekley & Ployhart, 2006). In a recent meta-analysis of the SJT literature, Christian, Edwards, and Bradley (2010) reported corrected criterion-related validity coefficients that ranged from .06 to .50, depending upon the construct assessed and the measure of job performance. SJTs also predict job performance beyond cognitive ability and the Big Five personality factors (McDaniel, Hartman, Whetzel, & Grubb, 2007), and racial subgroup differences for SJTs are much lower than those typically observed for cognitive ability tests (Whetzel, McDaniel, & Nguyen, 2008). This body of work indicates that SJTs are useful tools for predicting employee performance.

Little research has addressed the possibility that the validity of SJT-measured constructs may be influenced by other employee characteristics (cf. Chan, 2006), but such research is important for understanding the potential limits of SJTs. In summarizing the state of research on SJTs, Ployhart and Weekley (2006) wrote that “we must move beyond simply examining the criterion-related validity to examine the boundary conditions of operational SJTs” (p. 349). In this research, we examine boundaries of the criterion-related validity of SJTs. We posit that for some predictor constructs, the validity of inferences drawn from SJT scores may be dependent upon other traits. Specifically, for SJT-measured interpersonal skills, the applicants’ levels of *angry hostility* (AH) will influence how well the SJT predicts job performance. Interpersonal skills, also known as applied social skills (Christian et al., 2010; Huffcutt, Conway, Roth, & Stone, 2001), are defined as the skills and abilities that allow people to interact appropriately and effectively with others (Carpenter & Wisecarver, 2004; Gardner, 1983; Klein, DeRouin, & Salas, 2006; Sternberg, Conway, Ketron, & Bernstein, 1981). AH is defined as the tendency to experience anger, frustration and bitterness (Costa & McCrae, 1992). People high on AH tend to overreact to frustration, be less calm and relaxed, score lower on measures of likability and adjustment, and behave less consistently across situations (Costa & McCrae, 1995). We draw on the theory of planned behavior (Ajzen, 1991) to argue that having higher AH impairs a test taker’s prediction of his or her behavior in actual interpersonal work situations. Thus, we examine whether AH will disrupt the relation between SJTs measuring interpersonal skills (ISJT) scores and job performance.

Support for our hypothesis would have several important implications. From a practical standpoint, scores from ISJTs may only be valid for a portion of applicant test takers. This has important ramifications, considering the frequency of current and projected use of ISJTs by

practitioners. In a recent assessment of more than 250 companies' selection practices conducted by Development Dimensions International, 24% of staffing directors reported using SJTs, and 18% reported planning to use SJTs to their companies' practices in the future (Boatman & Erker, 2012). Moreover, in the Christian et al. (2010) meta-analysis, ISJTs comprised 20% of those SJTs saturated with a single construct. In addition, a sizable additional number of SJTs measured interpersonally oriented constructs such as teamwork, managing others, handling employee problems, and handling people. Thus, our study has implications for the usefulness of many SJTs, in that practitioners might need to consider AH in settings where ISJTs are used. Also, because ISJT validity depends on test takers accurately translating attitudes and intentions into future performance, our study provides a setting for an examination of the influence of AH on predicting future feelings (Loewenstein & Schkade, 1999) and translating intentions into behavior (Ajzen, 1991).

The remainder of this introduction proceeds as follows. We begin by framing ISJTs as measures of procedural knowledge about appropriate behaviors in work-related interpersonal situations (Motowidlo, Hooper, & Jackson, 2006). Next, we discuss the state of research on moderators of SJT validity. We then explain the rationale for AH as a moderator of ISJT validity.

### *The Nature of Situational Judgment Tests*

Motowidlo et al. (2006) defined SJTs as measures of procedural knowledge. When taking an SJT, applicants obtain higher scores when the options that they indicate as most appropriate or least appropriate closely match the beliefs about these behaviors provided by job experts during test development. Thus, to the extent an applicant's procedural knowledge is similar to that of subject matter experts (SMEs), applicants are likely to obtain higher SJT scores. Motowidlo et al. noted that "different SJTs can measure different types of procedural knowledge" (p. 60) and that some SJT items could be used to measure procedural knowledge about how to handle interpersonal situations. Recent research confirms the predictive validity of procedural knowledge in interpersonal situations (i.e., ISJTs). For example, Lievens and Sackett (2012) found that Belgian medical school applicants' ISJT scores predicted internship and job performance.

### *Moderators of Situational Judgment Test Validity*

To our knowledge, there is only one published study examining moderators of the SJT–performance relationship (Chan, 2006). Chan found that SJT scores, which he defined as an underlying construct *situational*

*judgment effectiveness* (SJE) or practical intelligence, moderated the relation between proactive personality and a number of work perceptions and work outcomes, including job performance. Proactive personality related positively to work variables among employees with high SJE and negatively among employees with low SJE. The results also show that proactive personality moderated the relation between SJE and outcomes, such that SJE was more strongly related to performance among those higher on proactive personality. Chan concluded that future researchers “need to proceed in a more theory-driven manner to hypothesize and test for moderators that affect the criterion-related validity of SJTs” (p. 480).

Our study builds on the work of Chan (2006) in two ways. First, we move toward a focus on theoretical moderators of specific predictor constructs by taking a construct-oriented view of SJTs. Thus, we address calls for understanding SJT validity in the context of the specific constructs measured (Arthur & Villado, 2008; Christian et al., 2010). Second, whereas Chan’s (2006) theorizing centered on proactive personality’s relation with performance (with SJE scores as moderators), our reasoning is focused on ISJT validity and AH as a moderator of this validity. This is consistent with our expectations of (a) a significant relationship between the ISJT and performance, and (b) a weak or nonsignificant relationship between AH and performance.

### *ISJTs and Moderators of the Intention–Behavior Linkage*

In this study, we focused only on SJT items measuring procedural knowledge in interpersonal skills-related situations. In order for ISJTs to predict interpersonally effective job behaviors, individuals have to consistently respond on the job in a manner similar to how they respond on the test. That is, people not only have to *know how to respond* to test items, they have to *enact the behaviors* reflected in those items when they are performing their jobs.

When responding to SJT items, test takers are essentially providing a prediction of their future behavior on the job. Unfortunately, there are many reasons why people are inaccurate in predicting their future behavior (Brooks & Highhouse, 2006). We argue that it is more difficult for some individuals than for others to predict, using an ISJT, how they will respond in interpersonal situations at work. Below, we develop two arguments for why those who have higher levels of trait AH will be particularly poor in their predictions.

*AH and the difficulty of translating intentions into behavior.* SJTs are given to individuals during a testing situation where they are not currently interacting with employees. As part of this process, SJT items require

individuals to judge how they will behave in a future situation that requires interpersonal interaction and thus may be more emotionally laden. Drawing from terminology adopted from the theory of planned behavior (Ajzen, 1991), responses on ISJTs could be characterized as either a statement of one's *attitude* toward a particular behavior (e.g., in should-do items, such as "setting up a meeting to state one's concerns to a difficult coworker about his behavior is the right thing to do") or of one's *intention* to engage in certain behaviors when one encounters the same or similar situations in the workplace (e.g., in would-do items, such as "in this situation, I would set up a meeting to state my concerns to a difficult coworker about his behavior"). In order for ISJT responses to predict job performance, employees must be able to translate their attitudes and intentions into future behavior at the time that the work environment demands it (e.g., actually choosing to set up a meeting with a difficult coworker).

Because of the nature of the testing environment, it is not expected that AH relates to ISJT performance. ISJTs are "low-fidelity simulations" (Motowidlo, Dunnette, & Carter, 1990) and provide fewer interpersonal contextual details than real work situations, which may be more emotionally laden. Test takers are merely expressing their procedural knowledge and indicating their *attitudes* toward a behavior or *intentions* to engage in future behavior. As such, AH is unlikely to affect responses to hypothetical interpersonal situations in a testing environment. However, as we explain, AH is a trait that is more strongly associated with behavior when provoked. Thus, AH should cause a disruption in the link between attitudes and future behavior, and between intentions and future behavior, thereby influencing the criterion-related validity of ISJT scores.

The theory of planned behavior (Ajzen, 1991; Ajzen & Fishbein, 1977) has focused on *behavioral control* as a key variable in the link between intention and behavior, such that when behavioral control is higher, the relationship between intentions and behavior should be stronger. Considerable research would lead one to predict that, even when people higher on AH may know what is appropriate and have intentions to act appropriately, they lack the behavioral control necessary to translate their intentions into behavior. In fact, research shows that people who are higher on trait AH have a sense of control that is not realistic (Lerner & Keltner, 2001). They have highly optimistic predictions of future events, "even when angry subjects rate the likelihood of events for which anger is a predisposing factor" (Lerner & Tiedens, 2006, p. 124). This means that, even though higher-AH people are more likely to experience a host of negative events—such as divorce, high blood pressure, and problems at work (Caspi, Elder, & Bem, 1987; Fredrickson et al., 2000)—they actually rate themselves as less likely than the average person to experience these events (Lerner & Keltner, 2001).

People who have higher levels of AH also have a number of other cognitive and social characteristics that could cause the intention–behavior link to be disrupted. For example, Berry, Worthington, O'Connor, Parrot, and Wade (2005) found that AH was related to vengeful rumination, or automatic and repetitive cognitions that prevent people from forgiving others for transgressions. Vengeful rumination may cause employees to be focused on coworkers' previous transgressions, such that they are unable to respond appropriately to them in future interpersonal interactions. Similarly, research also suggests that in terms of heart rate and blood pressure, people who are higher on AH take longer to recover physically from anger-inducing events (Fredrickson et al., 2000). This suggests that when difficult interpersonal situations occur at work, the effects of those events may linger for a longer period of time in the minds of those who are higher on trait AH and may prevent them from acting appropriately in interactions that occur shortly after the anger-inducing event. Finally, in laboratory simulations, those with higher levels of AH engaged in more inappropriate social behaviors (Murphy & Eisenberg, 1997).

Although we focused here on *trait* AH, the literature on *state* AH suggests a number of other reasons for the disruption of the intention–behavior link (e.g., Dunn & Schweitzer, 2005; Lerner & Tiedens, 2006; Tiedens & Linton, 2001). Although trait AH and state AH are not the same—state affect is more variable and shorter term (Tellegen, 1985)—they are correlated (Costa & McCrae, 1992) because personality-level traits can predispose a person toward experiencing state affect (Lazarus, 1991). Thus, findings in one area can provide insight about the other. Much of the literature has focused on the uniqueness of angry emotion, distinguishing it from other negatively valenced emotions, such as sadness and guilt. For example, Keltner, Ellsworth, and Edwards (1993) induced participants to feel either anger or sadness by having them imagine themselves in emotional situations. The high-anger group subsequently rated human causes of negative events as more likely (e.g., missing a flight because of a “terrible cab driver”), whereas the sadness group rated situational causes as more likely (e.g., missing a flight because of bad traffic). The authors concluded that anger leads to appraisals of events as being under the control of others, and this incidental anger carries over to unrelated situations (Lerner & Keltner, 2001). Another study using the same framework found that participants who were induced to feel anger subsequently rated coworkers as less trustworthy (Dunn & Schweitzer, 2005). These findings suggest that people who are relatively higher on trait AH may be more likely to assign blame and to distrust other people at work (e.g., Lerner & Tiedens, 2006).

Previous research is clear that AH, both as a trait and a state, is associated with a number of unique behaviors and cognitions we described above—overly optimistic future outlook, vengeful rumination, tantrums

and other inappropriate social behaviors, and a tendency to assign blame and distrust others. For these reasons, we predict that as compared to employees who are lower on AH, those who are higher on AH will be less likely to behave in a manner consistent with their responses to ISJT items. This leads to our expectation of weaker relations between ISJT scores and job performance among those high on AH.

*Angry hostility and the difficulty of predicting future feelings.* Another reason to expect an interaction between AH and ISJT scores is that individuals are not very good at predicting how they will feel in the future (Loewenstein & Schkade, 1999). This is especially true regarding their estimation of the effects of the situation on their future feelings. Research on such diverse areas as social influence (Milgram, 1965), protection against sexually transmitted disease (Gold, 1993), drug use (Lynch & Bonnie, 1994), and overspending on credit card purchases (Ausubel, 1991) suggest a consistent pattern of results: In the present, people tend to underestimate the influence of situations on their future behavior and to overestimate their self-control in future circumstances (Loewenstein, 1996; Loewenstein & Schkade, 1999). This research suggests that test takers may have a difficult time predicting how their feelings will affect their behavior in a "hot" state, such as when they are dealing with the pressure of the typical work day (Loewenstein, 1996). These misjudgments that occur between different visceral states have been termed "hot-cold empathy gaps" because people have a difficult time imagining what a different state will feel like (Loewenstein, 2000).

Previous research has shown that many people experience such difficulties; we posit that this difficulty is more pronounced for those who are higher on AH. Employees who are higher on trait AH are likely to have stronger reactions to difficult or negative interpersonal situations than will those lower in AH (Plutchik, 2003). For example, Judge, Scott, and Ilies (2006) found that trait AH moderated the relation between daily interpersonal injustice and (daily) state AH, such that the relation was stronger among those higher on trait AH. From a psychophysiological perspective, interpersonal stressors such as provocation (Suls & Wan, 1993) and disagreement about emotional issues (Davis, Matthews, & McGrath, 2000) led to greater increases blood pressure among those who were higher on trait AH. Though employees higher on trait AH may know the appropriate response in interpersonal situations, they are unable to consistently exert the self-control required to act appropriately (e.g., Bazerman, Tenbrunsel, & Wade-Benzoni, 1998).

In summary, given that ISJTs do not require actual interaction with other employees, it is likely that AH will only minimally influence responses to test items. Moreover, many workplace situations will not trigger AH to affect performance. Thus, we do not expect that AH will be negatively related to job performance. An important conclusion from

the Judge et al. (2006) and Davis et al. (2000) studies described above is that AH leads to stronger reactions in the presence of a trigger. Therefore, in workplace contexts where people have to exert restraint to enact appropriate interpersonal behaviors, trait AH is likely to prevent employees from exercising good judgment that is required for effective performance. Thus, the behavior of employees who have higher levels of AH is likely to be more *variable*. In some situations, they are able to use procedural knowledge about interpersonal behavior in an effective manner. However, in hot or emotional situations, their AH gets the best of them and they are less able to demonstrate effective interpersonal behavior. Thus, their performance is less predictable from their ISJT-measured procedural knowledge. Accordingly, we predicted:

*Hypothesis 1:* The relation between ISJT performance and job performance is moderated by AH, such that the positive relation between ISJT performance and job performance is stronger among those who have relatively lower levels of AH than those who have relatively higher levels of AH.

*Role of AH versus other facets of neuroticism.* Because AH is a facet of neuroticism, it is also important to consider whether any of the other five facets of neuroticism (i.e., depression, anxiety, self-consciousness, immoderation, and vulnerability to stress; Costa & McCrae, 1992) would moderate the ISJT–performance relationship in a way similar to the way AH does. Theory and empirical research, however, suggest that AH is unique and is distinct from other neuroticism facets in terms of the constellation of cognitions and behaviors associated with this trait. That is, the cognitive and social characteristics associated with AH that we discussed above are highly characteristic of AH and less so for the other facets of neuroticism. The same is true for the stronger and longer-lasting psychological and psychophysiological reactions to difficult or negative interpersonal situations (Fredrickson et al., 2000; Judge et al., 2006; Suls & Wan, 1993). Thus, AH is the facet of neuroticism that is relevant to our arguments. Nevertheless, it is also important to answer this question empirically. Across the three studies we present in this paper, we attempt to determine whether any of the other five facets of neuroticism function as a moderator of this relationship.

#### *Overview of the Three Studies*

In order to test our central hypothesis, we conducted three field studies with three main variables: ISJT scores, trait AH, and job performance.



Participants in Study 1 were employees in a healthcare organization, and the data were drawn from a concurrent validation study aimed at developing a battery of selection tests. In Study 2, participants were police officers; the data were drawn from a predictive validation study, in which employees completed the ISJT as part of a promotion process. Participants in Study 3 were applicants who took an ISJT as part of their medical school entrance exams.

### *Study 1*

#### *Method*

*Participants and Procedure.* Participants in Study 1 were 225 employees from a healthcare organization in the Northeastern United States. They held various job titles, including technician, associate, counselor, coach, assistant, and instructor. The sample involved individuals providing direct customer (patient) care. Provision of such care through interpersonal interactions with patients was the common element leading these jobs to be grouped together under a single selection system. Participants were drawn randomly from the incumbent population such that all incumbents had an equal chance of being selected for the study, with one exception in that low-tenured employees (e.g., less than one year) were not eligible for the study in order to limit the sample to employees for whom sufficient and stable performance information was available.

Demographic information was available for 67–74% of the sample, depending on the demographic variable. Of those who responded, 52% were female, with an average age of 35.3 years ( $SD = 10.8$ ). Forty percent of the participants were Black, 25% were White, 20% did not identify their race, 5% were Hispanic, 3% Native Hawaiian/Pacific Islander, 1% Native American, and 2% indicated a race other than those listed above. Of those who responded, 34% had a high school diploma or equivalent, 23% had an associate's or technical degree, 38% had a bachelor's degree, whereas 5% had a master's degree and less than 1% had earned a doctorate. Participants had an average organizational tenure of 2.68 years ( $SD = 2.31$  years) and position tenure of 2.21 years ( $SD = 1.43$  years).

The predictor scales were administered to incumbent employees during structured test administration sessions, as one stage of a criterion-related validation study conducted as the foundation for a redesign of the selection procedures for the target job family. Participants completed the predictor scales using a secure, Internet-based testing platform. Collection of competency-based performance data involved manager rating sessions, and these were preceded by frame-of-reference and rater error training. In

these sessions, supervisors evaluated employees using several behavioral statements for each competency. These competencies were established through a job analysis. Performance ratings were utilized for research purposes only and were described to participants as such.

### *Measures*

*Angry hostility and self-consciousness.* AH and self-consciousness (SC) were each measured with 10-item scales from the International Personality Item Pool Representation of the Revised NEO Personality Inventory (IPIP; Goldberg, 1999). As Schmitt (2008) noted, this measure is psychometrically comparable to the NEO-PI-R. An example AH item was, "I get irritated easily." Participants responded to the items on a five-point scale (1 = *strongly disagree*; 5 = *strongly agree*). The internal consistency reliability estimate ( $\alpha$ ) of this scale was .79. Self-consciousness ( $\alpha = .71$ ) was also measured with the IPIP-NEO. An example item was, "I am afraid to draw attention to myself."

*Interpersonal skills SJT.* The ISJT and the criterion measures were developed based on a job analysis, which was conducted as the first step in developing a test battery. An extensive job analysis was conducted, including a review of job-specific documentation, in-person job observations, business impact discussions with senior leaders, focus groups with 10 content experts to determine most important competencies, and questionnaires distributed to 114 job content experts. On the basis of this analysis, nine criterion competencies were developed.

To predict these criteria, the consultants developed a test battery that included an SJT, an action benchmarking response scale (e.g., rating the effectiveness of each action in a list for dealing with work problems), and a biographical data inventory. Given our research questions, we were interested only in the ISJT. The ISJT, including the items and the response options, were developed by the consulting firm in other healthcare settings. The initial development process involved critical incidents techniques in which incumbents and supervisors identified important workplace events and provided examples of highly effective and highly ineffective performance during those events. Consultants met with SME focus groups in the Study 1 organization to determine a final set of items that was most relevant to the specific jobs in this healthcare organization (i.e., those that most closely matched the competencies identified in the work analysis).

The instructions for each item were as follows: "Respond to the following situations by indicating the response you view as most appropriate and effective." As such, items were structured so that there was a single

correct response for each item, and the items were thus scored “0” for each incorrect response and “1” for each correct response.

*Construct validity of ISJT.* The original SJT had a total of 21 items, and the intention was to assess situational judgment in interpersonal, teamwork, and job-knowledge contexts. For the purpose of this study, two authors independently examined the original test and judged which of the items assessed interpersonal skills. One identified 16 items and the other identified 15 items as assessing interpersonal skills; after discussion, the 16th item was excluded.

We validated the interpersonal nature of these 15 ISJT items by asking five SMEs with PhDs in industrial-organizational psychology (80% male; mean years since earning PhD = 10.4) to indicate the extent to which each of the 15 items reflected two constructs (Christian et al., 2010). One was *interpersonal skills*, defined as *skills and abilities that allow people to interact appropriately and effectively with others*. The second was *job knowledge*, defined as *knowledge specific to a given field of work, including the application of appropriate occupational or organizational policies*. Each item was rated on the extent to which it tapped the content of each construct on a scale of 1 to 4 (1 = *not at all*; 2 = *to a small extent*; 3 = *to a moderate extent*; 4 = *to a large extent*). SMEs rated the ISJT items as tapping interpersonal skills to a moderately large extent ( $M = 3.56$ ,  $SD = .26$ ) and rated the ISJT items as tapping job knowledge to a small extent ( $M = 1.99$ ,  $SD = .20$ ). The difference between these means was significant ( $t = 10.39$ ,  $p < .01$ ). In addition, SMEs rated each of the focal items as more strongly tapping interpersonal skills content, compared to job knowledge content, with the exception of one item. As a result, we removed this one item from the ISJT, and the final measure contained 14 items. An example item is presented in the Appendix.

The 14-item scale had an estimated internal consistency reliability of  $\alpha = .41$ . Many authors have argued that internal consistency is not the optimal reliability estimate for use with SJT items (e.g., Lievens, Peeters, & Schollaert, 2008; Whetzel & McDaniel, 2009) and that alternate forms reliability and test-retest reliability are better measures. Studies of prior versions of the SJT content used in this study have produced test-retest reliability estimates that range between .60 and .70.

*Job performance.* Supervisors rated employee performance on each of the following competencies: adaptability (four items,  $\alpha = .95$ ), customer orientation (six items,  $\alpha = .96$ ), communication (five items,  $\alpha = .95$ ), continuous learning (six items,  $\alpha = .94$ ), teamwork (five items,  $\alpha = .95$ ), initiative (four items,  $\alpha = .94$ ), leveraging diversity (five items,  $\alpha = .93$ ), safety awareness (six items,  $\alpha = .95$ ), and work ethic (five items,  $\alpha = .93$ ). Across the 225 participants, 54 managers provided ratings, for an

average of 4.17 rates per rater. Each employee was rated by only one supervisor.

We examined the factorial validity of the criterion variables by performing a confirmatory factor analysis (CFA) in AMOS 20. Based on the information provided by the organization, we estimated a two-factor higher-order model. In this model, the indicators for each performance variable loaded on their respective latent variable factors. In turn, each of these latent variables loaded on one of two higher order factors: motivation-related performance (comprised of adaptability, continuous learning, initiating action, safety intervention, and work ethic) and interpersonal performance (comprised of the remaining performance variables). Although the chi-square value was significant,  $\chi^2$  (979) = 22046.57,  $p < .01$ , the comparative fit index (CFI = .92), standardized root mean residual (SRMR = .06), and the root mean square error of approximation (RMSEA = .07) suggested good fit of the model to the data (Hu & Bentler, 1999; Mathieu, Gilson, & Ruddy, 2006). However, inspection of the correlation between the two latent variables revealed a value of  $r = .98$ . Thus, we also estimated the model fit for a one-higher-order factor model. In this analysis, the hypothesized model also fit the data well (CFI = .92, SRMR = .06, RMSEA = .07). Given the good fit of the one-higher-order factor model and our desire for parsimony, we treated the performance ratings as reflecting a unidimensional variable ( $\alpha = .99$ ).

### *Results and Discussion*

Table 1 presents the means, standard deviations, and intercorrelations among the study variables. The correlation between the ISJT measure and the performance criterion was .19 ( $p = .01$ ).

To test the hypothesis, we conducted a moderated regression analysis. Because none of the correlations between demographic variables and the variables of interest was significant, we elected not to control for demographics, in order to preserve degrees of freedom. We entered SC, AH, and ISJT score on the first step, the SC  $\times$  ISJT interaction on the second step, and the AH  $\times$  ISJT term on the third step.

The results of this analysis are presented in Table 2. Inspection of this table reveals initial support for the hypothesis.<sup>1</sup> The AH  $\times$  ISJT interaction explained significant incremental variance in the outcome,

---

<sup>1</sup>Although it is possible that angry hostility is a suppressor variable that could account for increases in the predictive validity of the ISJT, the data suggest that this is not the case. The variance explained by the ISJT by itself (i.e., the squared bivariate correlation) did not increase meaningfully when the personality variables were entered in the same step.

TABLE 1  
*Means, Standard Deviations, and Intercorrelations Among Study 1 Variables*

Variable	Mean	SD	1	2	3	4	5	6	7	8	9
1. Gender	1.52	.50	–								
2. Age	35.27	10.80	–.05	–							
3. Education	3.15	.97	.05	.08	–						
4. Organizational tenure	2.68	2.31	–.11	.36*	–.01	–					
5. Position tenure	2.21	1.42	–.09	.43*	–.12	.50*	–				
6. Angry hostility	1.74	.54	.08	–.13	.03	–.08	.06	(.79)			
7. Self-consciousness	1.94	.52	.25*	–.08	.03	–.07	–.05	.41*	(.71)		
8. ISJT scores	9.57	2.02	.17*	–.02	–.07	.06	–.03	–.14*	–.04	(.41)	
9. Performance	3.42	.86	–.02	–.10	–.07	–.03	–.12	–.09	.03	.19*	(.99)

*Note.* Because of missing data for demographic variables, *N* for correlations that include those variables varies from 142 to 225. ISJT = interpersonal skills situational judgment test. Decimals in correlations omitted for clarity. Reliabilities presented in parentheses on the diagonal. Gender is coded 1 = male, 2 = female; Education is coded 1 = some high school, 2 = high school diploma or equivalent, 3 = two-year or associate's degree, 4 = four-year undergraduate degree, 5 = master's degree, 6 = doctorate.

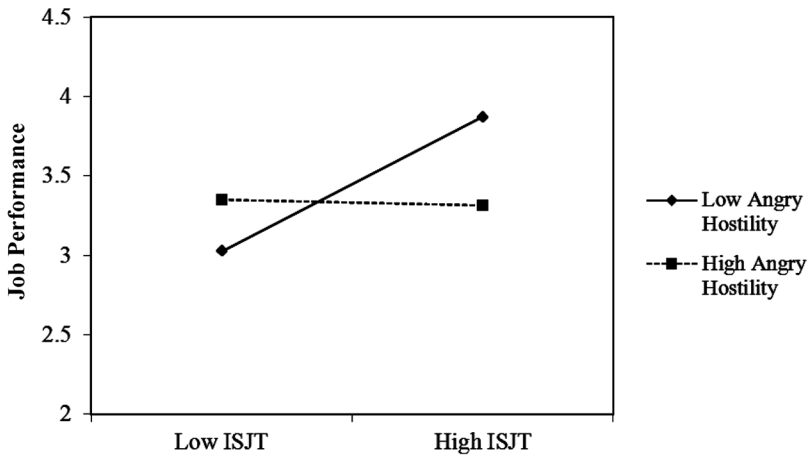
\* $p < .05$ .

TABLE 2  
*Moderated Regression Analysis for Study 1*

Predictor	<i>B</i>
Step 1	
Self-consciousness (SC)	.06
Angry hostility (AH)	–.08
ISJT	.15*
$\Delta R^2$	.04*
Step 2	
SC $\times$ ISJT	–.06
$\Delta R^2$	.01
Step 3	
SC $\times$ ISJT	.03
AH $\times$ ISJT	–.22**
$\Delta R^2$	.06**
$R^2$	.11**

*Note.*  $N = 225$ . ISJT = interpersonal skills situational judgment test. Step 1: Add SC, AH, and ISJT; Step 2: Add SC  $\times$  ISJT; Step 3: Add AH  $\times$  ISJT. All regression coefficients reported in this table are unstandardized (*B*). The results for both interaction terms (SC  $\times$  ISJT; AH  $\times$  ISJT) are reported under the heading "Step 3" to facilitate direct comparisons of the interaction terms.

\* $p < .05$ . \*\* $p < .01$ .



**Figure 1: Interaction of ISJT Score and Angry Hostility Predicting Overall Job Performance (Study 1).**

*Note.* ISJT = interpersonal skills situational judgment test. High = 1 SD above the mean; Low = 1 SD below the mean.

after controlling for the main effects, SC, and the SC  $\times$  ISJT interaction. Specifically, the interaction term accounted for 6% of the variance in job performance. The negative value for the coefficient for the interaction term suggests that the direction of the interactions was as predicted. Given the somewhat modest sample size, we also utilized bootstrapping procedures with 1,000 resamples in SPSS 20.0, in order to calculate 95% confidence intervals (CIs) around the estimates of each of the interactions terms. Results confirmed that the interaction was significant (95% CI for  $B = -.31, -.10$ ).

To more explicitly test our hypothesis, we graphed the interactions, following the procedures outlined by Aiken and West (1991). We plotted the regression lines at  $-1$  SD and  $+1$  SD for ISJT score and for the low and high AH groups (also at  $-1$  SD and  $+1$  SD). Figure 1 shows the AH  $\times$  ISJT interaction predicting overall performance. Whereas the line is virtually flat for those who are higher on self-rated AH, the relationship is strong and positive for those who are lower on self-rated AH. These results were confirmed with analyses of simple slopes, regressing criterion variables on ISJT at  $-1$  SD and  $+1$  SD for AH. The slope at  $-1$  SD for AH was  $b = .42, p < .01$ ; at  $+1$  SD AH, the slope was  $b = -.02, p = .84$ .

It is also worthwhile exploring the differential validity of the ISJT for the low-AH and high-AH groups. In order to examine this, we split the sample at the median of AH and correlated ISJT scores with criterion

scores for each group. Despite the fact that the overall criterion-related validity was significant ( $r = .19$ ), the ISJT was unrelated to performance among those high in AH ( $r = .00$ ,  $p = 1.00$ ). In contrast, in the low-AH group, the relation between the ISJT and performance was positive, strong, and significant ( $r = .43$ ,  $p < .01$ ). Again, these results provide support for the study's hypothesis.

### *Study 2*

In Study 2, we sought to replicate the findings in a very different job from the focal job in Study 1—a sample of officers from an urban police department. Although interpersonal skills are important for many jobs, the work activities requiring those skills vary greatly across settings. For example, the O\*NET indicates high levels of importance for interpersonal skills for both police officers and medical assistants. However, police work focuses much more on resolving conflicts or negotiating with others, whereas for medical assistants, greater importance is placed on assisting and caring for others (Mumford & Peterson, 1999).

Study 1 used a concurrent validation process. In contrast, the ISJT in Study 2 was part of an operational promotion process, where ISJT scores (along with other assessments) were used to make promotion decisions. Given the high-stakes nature of the testing, Study 2 represents a stronger test of the hypothesis. Our hypothesis rests on the condition that AH is weakly related to ISJT scores. This is quite likely in low-stakes testing (e.g., with concurrent validation studies) because test takers feel less pressure. However, the increased tension that could be caused by the high-stakes testing situation could trigger more emotional reactions among those higher on AH, which could lead to a stronger relationship between AH and ISJT scores.

### *Method*

*Participants and Procedure.* Participants in Study 2 were police officers in a metropolitan area in the Southwestern United States. They were either officers applying for a promotion to sergeant or sergeants applying for promotion to lieutenant. Demographic information was available for 95–97% of the sample. The sample was 89.8% male with an average age of 38.3 years ( $SD = 6.3$ ). Sixty-nine percent of the officers were White; 21% were Hispanic, 4% did not identify their race; 1% were Black; 1% Native American, and 4% indicated a race other than those listed above. Five percent had a high school diploma or equivalent, 11% had an associate's or technical degree, 38% had some college, 30% completed a bachelor's degree,

4% had completed some graduate school, and 8% had a master's degree or a doctorate. Officers had worked at the organization for an average of 13.79 years ( $SD = 4.39$  years).

As noted above, the Study 2 sample involved individuals participating in an operational promotional assessment that was being conducted by the police department and city human resources. All 188 officers who were part of the promotion process were invited to participate; from this group, 188 took the ISJT, 98 (52%) took the Internet-based personality survey, and we obtained supervisor performance ratings for 54 (29%). Officers who participated in all three phases (i.e., their performance was rated by their supervisors) did not differ from those who only completed the personality survey on demographic (age, racial mix, gender mix education) or key study variables (ISJT, neuroticism, self-rated interpersonal skills). When comparing those who only took the ISJT and those who completed the personality survey, the latter group had a significantly higher level of education ( $M = 3.4$  vs.  $M = 2.75$ ),  $F(1, 121) = 6.16, p = .01$ , although the modal education level for both groups was "some college."

The trait survey and performance ratings occurred approximately two years after the promotional exam. Participants completed the individual-differences survey using a secure, Internet-based testing platform; supervisor ratings of officer performance used a similar platform. The supervisor of each officer who participated in the trait survey was asked to rate the officer using several behavioral statements for each performance dimension. Performance ratings were utilized for research purposes only and were described to participants as such. Participants in the supervisor rating survey were the current supervisor for each ratee to which they were assigned.

Although the operational assessment included a job knowledge test, an assessment center, and an oral board review, we did not consider this information as it was not relevant to our research focus. Promotions were conditional on the aggregate performance on all of these measures, so the ISJT score was not the sole determinant of promotion. The promotion decision was not a source of range restriction in this study as the great majority of officers who took the test but who were not promoted remained with the department, and thus they were available for all further data collection efforts.

### *Measures*

*Facets of neuroticism.* AH ( $\alpha = .83$ ) was measured with the same 10-item scale as was used in Study 1 from the IPIP-NEO (Goldberg, 1999). In order to determine whether other facets of neuroticism might also



interact with ISJT scores, we also measured anxiety ( $\alpha = .76$ ), depression ( $\alpha = .80$ ), vulnerability to stress ( $\alpha = .78$ ), and self-consciousness ( $\alpha = .84$ ) using the IPIP-NEO.<sup>2</sup> Each of these facets was also measured with 10 items, using a five-point scale of agreement. Example items were “Worry about things” (anxiety); “Often feel blue” (depression); “Become overwhelmed by events” (vulnerability to stress); and “Am afraid I will do the wrong thing” (self-consciousness).

*Self-rated interpersonal skills.* To establish the construct validity of the ISJT, we measured interpersonal skills using six items from Ferris, Witt, and Hochwarter (2001), using a five-point scale (1 = *strongly disagree*; 5 = *strongly agree*). Sample items include, “I am keenly aware of how I am perceived by others,” “In social situations, it is always clear to me exactly what to say and do,” and “I am able to adjust my behavior and become the type of person dictated by any situation.” The internal consistency reliability estimate ( $\alpha$ ) of this scale was .73.

*Supervisor-rated interpersonal skills.* We also had supervisors rate the interpersonal skills of each employee to establish the construct validity of the ISJT. Supervisors rated employees using a five-point scale (1 = *never*; 5 = *always*) on 13 items such as “Demonstrates interpersonal skills when on the job.” The internal consistency reliability estimate ( $\alpha$ ) was .81.

*ISJT.* The ISJT was developed using the critical incident technique. The first step was the selection of an SME committee that was tasked with evaluating the core competencies for the position of sergeant or lieutenant (e.g., decision making, personal responsibility, conflict management, political savvy, perception of social cues). The SME committee then developed scenarios based on real-life situations that reflected the core competencies. These scenarios were then broken down into series of *would-do* responses scored as *most effective* to *least effective*. For the purpose of this study, the first and second authors returned to the item content to determine which items were saturated with interpersonal skills. We used a process that was similar to Study 1, whereby items were classified by the predominant construct that they tapped. Items were considered “saturated” with interpersonal skills when the researchers determined interpersonal skills to be the predominant construct tapped by the item (as opposed to job knowledge or other constructs). After initial independent classification of the items, there were two disagreements, which were then resolved by discussion. After this resolution, of the 30 SJT items for each

---

<sup>2</sup>Note that we test these four neuroticism facets, but we did not test immoderation in Study 2. Based on discussion with our contact at the Study 2 organization, we decided not to measure immoderation because so many of the items in the IPIP-immoderation scale are food-binge-related (e.g., “Love to eat,” “Go on binges,” “Often eat too much”) and not relevant to the current context.

test, six were saturated with interpersonal skills for the sergeants' SJT, and seven were saturated with interpersonal skills for the lieutenants' SJT. As the item content varied for the sergeants' and lieutenants' processes, we computed separate scores for each, standardized them (within-SJT), and combined them into a single variable. An example item is presented in the Appendix.

*Task performance.* Supervisors rated employee in-role performance using a five-point scale (1 = *never*; 5 = *always*) using four items from Williams and Anderson (1991). These items were "Adequately completes assigned duties," "Fulfills responsibilities specified in the job description," "Performs tasks that are expected of him/her," and "Meets formal performance requirements of the job." The internal consistency reliability estimate ( $\alpha$ ) of this scale was .79.

*Contextual performance.* Supervisors rated three dimensions of employee contextual performance, using a five-point scale (1 = *never*; 5 = *always*). They rated their subordinates on helping (five items, adapted from Van Dyne & LePine, 1998); voice (four items, Van Dyne & LePine, 1998); and loyalty (three items, Moorman & Blakely, 1995). Example items were "volunteers to help coworkers without being asked" (helping); "provides constructive suggestions about ways to improve the unit's effectiveness" (voice); and "talks positively about the department to others" (loyalty). Internal consistencies for each dimension were helping ( $\alpha = .85$ ), voice ( $\alpha = .90$ ), and loyalty ( $\alpha = .85$ ).

*Performance dimensionality.* We conducted a CFA on these 16 items, and found that a four-factor model (separate factors for task performance, helping, voice, and loyalty) provided reasonable fit to the data, although the fit statistics were not particularly strong:  $\chi^2$  (113) = 197.92,  $p < .01$ , CFI = .85, SRMR = .10, RMSEA = .12. Of course, this is likely in part due to the small sample size relative to the number of indicators ( $N = 54$  with 16 indicators). Nevertheless, the four-factor model did fit the data better than a two-factor model in which all of the OCB items loaded on the same factor,  $\Delta\chi^2$  (5) = 127.92,  $p < .01$  and better than a one-factor model, in which all performance items loaded on the same latent factor,  $\Delta\chi^2$  (6) = 210.24,  $p < .01$ . Therefore, we retained each performance variable as a separate scale.

### *Results and Discussion*

Table 3 presents the means, standard deviations, and intercorrelations among the study variables. It is notable that AH was not related to any of the outcome variables. Correlations between the ISJT measure and the four performance criteria were  $r = .24$  ( $p = .08$ ) for task performance,

TABLE 3  
Means, Standard Deviations, and Intercorrelations Among Study 2 Variables

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1. Gender	1.07	.29	—																
2. Age	38.33	6.32	11	—															
3. Tenure	13.79	4.39	06	62*	—														
4. Education	3.42	1.18	−03	16	05	—													
5. Anxiety	2.34	.56	19†	−00	05	07	(76)												
6. Depression	1.38	.45	14	07	−02	−17	50*	(80)											
7. Vulnerability	1.41	.38	19†	14	05	−00	38*	47*	(78)										
8. Self-cons.	1.96	.63	06	00	06	−12	52*	52*	46*	(84)									
9. Angry hostility	2.19	.61	−03	31*	17	−01	43*	33*	30*	39*	(83)								
10. Prom. decision	.25	.44	03	−07	12	18†	11	−08	18†	15	03	—							
11. Self interp. skills	3.58	.52	−12	−28*	−18†	02	−05	−22*	−29*	−36*	−32*	−08	(73)						
12. Super interp. skills	3.63	.53	−10	−01	02	14	−02	16	23*	−03	−13	20	28*	(81)					
13. ISJT	.00	.99	−03	00	16	03	11	−07	05	−06	01	34*	23*	31*	(30)				
14. Task performance	4.69	.37	−12	12	05	05	04	15	09	−15	11	−06	24†	38*	24†	(86)			
15. Voice	3.51	.81	−15	10	19	−05	−09	−10	06	−11	06	−14	−07	03	31*	16	(90)		
16. Helping	3.86	.59	06	−00	−05	−09	11	−05	30*	12	12	04	10	24†	27*	25†	50*	(85)	
17. Loyalty	3.40	.80	−16	21	21	−08	04	−12	20	08	02	26†	04	15	22	22	17	34*	(85)

Note. For gender  $N = 95$ ; age  $N = 73$ ; tenure  $N = 93$ ; education = 95; promotion decision  $N = 98$ ; anger  $N = 96$ ; interpersonal skills  $N = 97$ ; ISJT  $N = 98$ ; Task performance, voice, helping, loyalty  $N = 54$ . Decimals in correlations omitted for clarity. Reliabilities presented in parentheses on the diagonal. Vulnerability = vulnerability to stress; Self-cons. = self-consciousness; Prom. decision = promotion decision; Self interp. skills = self-rated interpersonal skills; Super interp. skills = supervisor-rated interpersonal skills. Gender is coded 1 = male, 2 = female; Education is coded 1 = some high school, 2 = high school diploma or equivalent, 3 = two-year or associate's degree, 4 = four-year undergraduate degree, 5 = master's degree, 6 = doctorate; Promotion decision is coded 1 = promoted; 0 = not promoted. ISJT scores are standardized.

† $p < .10$ , \* $p < .05$ .

TABLE 4  
*Summary of Moderated Regression Analyses for Study 2*

Predictor	Task performance	Voice	Helping	Loyalty
Step 1				
AH	.08	.09	.12	.04
ISJT	.09 <sup>†</sup>	.26*	.17*	.18
$\Delta R^2$	.07	.10 <sup>†</sup>	.09 <sup>†</sup>	.05
Step 2				
AH $\times$ ISJT	-.20*	.00	-.21	-.23
$\Delta R^2$	.07*	.00	.03	.02
$R^2$	.14	.10	.12	.07

Note.  $N = 54$  for each analysis. ISJT = interpersonal skills situational judgment test; AH = angry hostility. All regression coefficients reported in this table are unstandardized ( $B$ ).

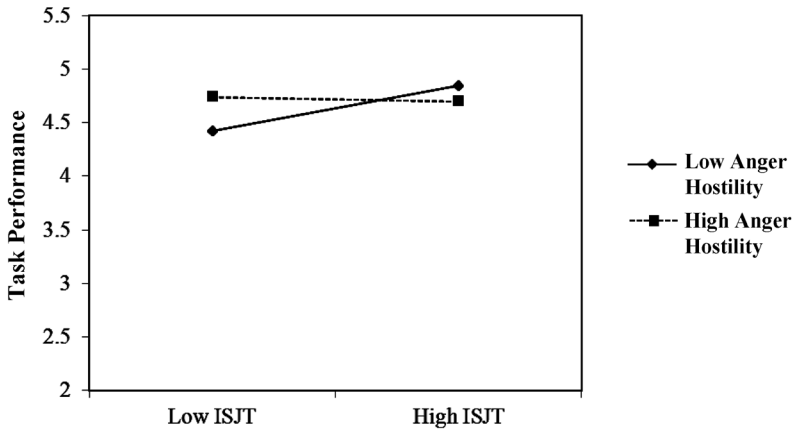
<sup>†</sup> $p < .10$ . \* $p < .05$ .

$r = .31$  ( $p = .02$ ) for voice,  $r = .27$  ( $p = .05$ ) for helping, and  $r = .22$  ( $p = .11$ ) for loyalty. As expected, whether an officer was promoted related significantly to ISJT scores ( $r = .34$ ,  $p = .01$ ).

**Construct Validity of ISJT.** We ascertained the degree to which the ISJT measured interpersonal skills by examining (a) its correlation with self-rated interpersonal skills and (b) its correlation with supervisor ratings of the interpersonal skills of each participant. As shown in Table 3, the ISJT score was positively correlated with self-rated ( $r = .23$ ,  $p = .02$ ) and supervisory-rated ( $r = .31$ ,  $p = .02$ ) interpersonal skills. The correlation between self- and other-rated interpersonal skills was also significant ( $r = .28$ ,  $p = .04$ ).

**Hypothesis Tests.** To test our central hypothesis, we conducted separate moderated regression analyses for each job performance variable. Because none of the correlations between demographic variables and variables of interest was significant, and because of our small sample size, we elected not to control for demographics in order to preserve degrees of freedom, and we conducted a separate regression analysis for each facet of neuroticism. In each analysis, we entered the relevant facet (self-consciousness, anxiety, depression, vulnerability, or AH), ISJT score on the first step, and the interaction term (e.g., AH  $\times$  ISJT) on the second step.

The results of the analyses for AH are presented in Table 4. Inspection of Table 4 reveals that for task performance, the interaction between AH  $\times$  ISJT interaction explained significant incremental variance in the outcome, after controlling for the main effects. The interaction term accounted for 7% of the variance in task performance. The interaction term was not significant for any contextual performance variables. However, the interaction term did account for 3% of the incremental variance in



**Figure 2: Interaction of ISJT Score and Angry Hostility Predicting Supervisor-Rated Task Performance (Study 2).**

*Note.* ISJT = interpersonal skills situational judgment test. High = 1 SD above the mean; Low = 1 SD below the mean.

**TABLE 5**  
*Correlations Between ISJT Scores and Supervisor Ratings by Anger Group, Study 2*

	Full data set ( <i>N</i> = 54)	Low-AH ( <i>n</i> = 23)	High-AH ( <i>n</i> = 31)
Task performance	.24 <sup>†</sup>	.51 <sup>*</sup>	.01
Voice	.27 <sup>*</sup>	.20	.37 <sup>*</sup>
Helping	.31 <sup>*</sup>	.44 <sup>*</sup>	.14
Loyalty	.22	.26	.18

*Note.* Low- and high-AH groups were created using a median split. AH = angry hostility.  
<sup>†</sup> $p < .10$ . <sup>\*</sup> $p < .05$ .

helping behavior and 2% of the incremental variance in loyalty. Although the interaction term was only significant for task performance, the negative values for the regression weights for helping and loyalty indicate that the direction of the interactions were as predicted. As with Study 1, we also utilized bootstrapping procedures and found that the 95% CI for  $B = -.42, -.03$ , suggesting additional support. The graph of the interaction (see Figure 2) and test of simple slopes ( $b = .22, t = 2.75, p = .01$  when AH was low;  $b = -.02, t = -.24, p = .81$  when AH was high) were also supportive of the hypothesis.

Table 5 presents differential validity of the ISJT for the low-AH and high-AH groups, and reveals an interesting pattern of results. For task

performance, helping, and loyalty, the results were similar, where the relation between ISJTs and performance was stronger for those low in AH (though again, this was only significant for task performance). However, for voice, the relation between ISJTs and voice was significant among those high in AH but not those low in AH. Tentatively, this suggests that ISJTs are weaker predictors of voice behavior among those who tend to have weaker emotional reactions to work events.

*Interactions With Other Facets of Neuroticism.* As discussed above, we also sought to determine whether any of the other neuroticism facets we measured—self-consciousness, anxiety, depression, or vulnerability to stress—interacted with ISJTs in a manner similar to AH. To check this, we conducted an additional 16 moderated regression analyses, where the interactions between ISJT and (a) self-consciousness, (b) anxiety, (c) depression, and (d) vulnerability to stress were used to predict (a) task performance, (b) voice, (c) helping, and (d) loyalty. Of these 16 analyses, the interaction was significant in only one<sup>3</sup>: The relationship between self-consciousness (SC) and ISJT scores was significant in predicting task performance, after controlling for the main effects of SC and ISJT,  $B = -.24$ ,  $\Delta R^2 = .10$ ,  $p = .01$ . Bootstrap analyses confirmed the significance of the interaction term, (95% CI for  $B = -.42, -.04$ ). Simple slopes tests also showed that the form of the  $SC \times ISJT$  interaction was similar to the  $AH \times ISJT$  interaction. For those low on SC, the simple slope for ISJT predicting task performance was positive and significant,  $b = .20$ ,  $t = 3.07$ ,  $p < .01$ ; for those high on SC, the slope was not significant,  $b = -.10$ ,  $t = -1.08$ ,  $p = .29$ .

The results of Study 2 supported the hypothesis that AH moderates the relation between ISJT scores and job performance, such that the relation between ISJTs and performance was strong and significant for those low on AH, but ISJTs did not predict performance for those high on AH. However, this was only true for task performance.

The follow-up analyses showed that neither depression nor anxiety nor vulnerability to stress interacted with ISJTs. However, self-consciousness did act as moderator, in the same way that AH did: The validity of the ISJT was strong when SC was low ( $r = .50$ ,  $p = .01$ ) but not when SC was high ( $r = -.09$ ,  $p = .68$ ). This is an interesting and unexpected finding because the  $SC \times ISJT$  interaction was not significant in Study 1. When considered across the two studies, this suggests that although the  $AH \times ISJT$  interaction generalizes across contexts, the  $SC \times ISJT$  interaction may be context-specific. For example, it is possible that this interaction surfaces in the context of police work because in this particular job there

---

<sup>3</sup>Given space limitations, we do not present the remaining nonsignificant results here; complete results are available from the first author.

are many interpersonal contexts that require officers to act in a way that is tactful but also forceful (e.g., apprehending violent or belligerent suspects). In most work contexts, where *inhibition of behavior* is required for effective interpersonal performance, AH prevents employees from consistently converting their good intentions into behavior. However, in some contexts, where *initiation of behavior* is required for effective interpersonal performance, SC may also function as a moderator.

### *Study 3*

As discussed above, we conducted Study 2 in order to replicate the findings in a completely different type of job from Study 1, but one that also had high levels of requirements for interpersonal skills in order to perform well. Given the importance of replication for understanding new phenomena in the social sciences (Schmidt, 2009; Tsang & Kwan, 1999), we conducted a third study. The study was in a true selection context, involving a sample of students who had taken the SJT as part of their requirements for entry into medical school. For the third study we also used an SJT that was specifically designed to measure interpersonal skills and had been validated as such (Lievens, Buyse, & Sackett, 2005a; Lievens & Sackett, 2012). This SJT was video based, which should provide increased generalizability. Finally, as we discuss below, in this study we also wanted to determine whether any of the five other facets of neuroticism interacted with ISJT to influence performance. We were particularly interested in the SC  $\times$  ISJT interaction, as this was significant in Study 2 but not in Study 1.

### *Method*

*Participants and procedure.* The initial sample of participants in Study 3 was a group of 941 students who had attended an admission exam for medical studies in Belgium. This sample was 61.8% female with an average age of 18.3 years ( $SD = 1.7$  years). More than 99% of this sample was White. The exam consisted of a cognitive part (general cognitive ability test) and a noncognitive part (video-based SJT, see Lievens, Buyse, & Sackett, 2005a, for details on the exam). Personality data were gathered during classes at the beginning of the first academic year in each of the medical universities in Belgium. Therefore, only students who had passed the admission exam and had entered the first year of medical studies were included. During first-year classes, students were asked to complete a personality inventory and to grant access to their academic records throughout their medical school career. Participants were informed that

their results were only to be used for research purposes and guaranteed confidentiality of the information provided. Specifically, 530 students (65% female, 99.5% Caucasian) with an average age of 18.2 years ( $SD = 1.4$  years) completed the personality inventory. There were no significant differences between this group and the population of students entering medical studies. Although two other publications (Lievens, Coetsier, De Fruyt, & De Maeseneer, 2002; Lievens, Ones, & Dilchert, 2009) have used the personality scores from these data, our study has minimal overlap because (a) we focus on ISJT scores and the other studies do not, and (b) we collected unique criterion data as we describe below.

### Measures

*Personality measure.* Participants completed the authorized Flemish translation (Hoekstra, Ormel, & De Fruyt, 1996) of the *NEO Personality Inventory—Revised* (NEO PI-R; Costa & McCrae, 1992). The NEO PI-R is a 240-item personality inventory assessing the Big Five dimensions of neuroticism, extraversion, openness, agreeableness, and conscientiousness, as well as six specific facets per factor (eight items per facet). Item scores were summed to arrive at overall facet scores. Previous research with the translated scale suggests evidence of good reliability and validity. For example, Hoekstra et al. reported neuroticism-facet alpha reliability estimates that ranged from .74 (impulsiveness) to .88 (AH), and averaged .80 (for comparison purposes, the original, English version of the NEO PI-R had an average facet reliability of .75). The item response scale ranges from 1 = *strongly agree* to 5 = *strongly disagree*. In the context of this study's hypotheses, only the neuroticism facets were used.

*ISJT.* The general aim of the video-based SJT was to measure interpersonal skills (see Appendix for an example item). To develop the video-based SJT, realistic critical incidents were collected regarding two key interpersonal domains ("building and maintaining relationships" and "communication/exchanging information"; see Carpenter & Wisecarver, 2004 Klein et al., 2006) from experienced physicians and professors in general medicine. Second, test developers wrote vignettes that included these incidents. Two professors teaching physicians' consulting practices tested these vignettes for realism. Using a similar approach, questions and response options were derived. Third, semiprofessional actors were hired and videotaped. Finally, a panel of experienced physicians and professors in general medicine developed a scoring key. Agreement among the experts was satisfactory, and discrepancies were resolved upon



discussion. The scoring key indicated which response alternative was correct for each item (+1 point). In its final version, the SJT consisted of videotaped vignettes of interpersonal situations that physicians are likely to encounter with patients. After each critical incident, the scene froze, and candidates received 25 seconds to answer the question ("What is the most effective response?"). In total, the SJT consisted of 30 multiple-choice questions with four possible answers. Prior research showed that scores on the ISJT consistently correlated with scores on interpersonally oriented courses (Lievens et al., 2005a).

In terms of reliability, the estimated internal consistency reliability ( $\alpha$ ) was .44. However, prior studies revealed that the alternate form reliability of the SJTs was .66 (Lievens, Buyse, & Sackett, 2005b), which is consistent with values obtained in studies on alternate-form SJTs (Catano, Brochu, & Lamerson, 2012; Clause, Mullins, Nee, Pulakos, & Schmitt, 1998).

*Job performance rating.* In light of this study's hypotheses, it was important to examine the validity of the ISJT against an interpersonally oriented criterion. Therefore, we used the job performance rating that students received when they worked under the supervision of a registered general practitioner in a general practice placement. These criterion data were gathered nine years after completion of the admission exam. In fact, of the students of this cohort who completed their seven years of education ( $N = 314$ ), about 30% ( $N = 92$ ; 65% female, mean age = 18.1 years [at the time of the original testing],  $SD$  age = .7 years) chose a career in general medicine and entered a general practitioner training program of up to two years duration. During that program, they worked under the supervision of a registered general practitioner in a number of general practice placements. Although being fully responsible for patients, they were evaluated on their technical (e.g., examination skills) and interpersonal skills (e.g., contact with patients) using detailed score forms. General practitioners were rated on various dimensions and on overall job performance. Slightly different forms were used for rating general practitioners' job performance, meaning that some forms consisted of fewer dimensions to be rated. Only the global job performance ratings (0 to 20) were available from the archives, and we standardized these overall job performance ratings (within-form) because the difference in forms may have led to different weighting of the dimensions in providing a general job performance rating. Note that none of the supervisors had access to the trainees' admissions scores. In terms of construct-related validity, prior research showed that these global job performance ratings were moderately (corrected  $r = .40$ ) correlated with internship ratings gathered during the medical studies (Lievens & Sackett, 2012).

TABLE 6  
*Means, Standard Deviations, and Intercorrelations Among Study 3 Variables*

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10
1. Gender	1.62	.49	—									
2. Age	18.16	1.50	.14*	—								
3. Angry hostility	21.00	4.68	-.03	.08	—							
4. Anxiety	24.69	5.69	.15*	.11*	.37*	—						
5. Depression	23.94	5.31	.15*	.07	.42*	.72*	—					
6. Self-consciousness	23.87	4.93	.14*	.11*	.36*	.57*	.57*	—				
7. Impulsiveness	25.41	4.75	.16*	.03	.37*	.21*	.28*	.19*	—			
8. Vulnerability	19.75	4.93	.13*	.08	.36*	.68*	.59*	.50*	.30	—		
9. ISJT	25.83	3.13	.11*	.05	.04	.07	.07	.07	.02	.06	(.44)	—
10. Job performance	.03	1.00	.02	.10	-.04	.06	.03	-.06	.04	.02	.23*	—

*Note.* For demographic variables and ISJT,  $n = 941$ . For personality variables,  $N = 530$ . For job performance,  $N = 92$ . ISJT = interpersonal skills situational judgment test. Decimals in correlations omitted for clarity. Gender is coded 1 = male, 2 = female.

\* $p < .05$ .

*Results and Discussion*

Means, standard deviations, and intercorrelations among Study 3 variables are presented in Table 6. Note that the ISJT was significantly related to job performance ( $r = .23$ ,  $p = .03$ ). Note also that the neuroticism facets were unrelated to the ISJT and to job performance (all  $p > .05$ ).

Given the multicollinearity among the facets of neuroticism (mean  $r = .43$ , with 7 of the 15 correlations at  $r = .50$  or higher), we conducted moderated regression analyses for each of the six facets of neuroticism to test our hypothesis and rule out the other facets as drivers of the moderating relationship. In each case, we entered the ISJT on the first step, the neuroticism facet on the second step (e.g., AH), and the interaction term (e.g.,  $AH \times ISJT$ ) on the third step.

Table 7 presents the results. Inspection of this table reveals that none of the neuroticism facets was related to job performance in Step 2, as expected from the observed nonsignificant correlations in Table 5. Table 7 also shows that, of the six interactions tested, only the  $AH \times ISJT$  interaction was significant,  $B = -.01$ ,  $t = -2.14$ ,  $p = .04$ , and this interaction explained 5% of the incremental variance in job performance ratings. The negative value of the coefficient suggests support for the hypothesis. Bootstrapping with 1,000 resamples confirmed that the interaction was significant (95% CI for  $B = -.026$ ,  $-.001$ ). Figure 3 shows the  $AH \times ISJT$  interaction predicting task performance and shows that the simple slope for ISJT predicting performance was significant ( $b = 2.69$ ,  $t = 2.37$ ,  $p = .02$ ) when AH was low; when AH was high, the simple slope was actually significant

TABLE 7  
*Moderated Regressions for Six Facets of Neuroticism Interacting With ISJT to Predict Performance (Study 3)*

	AH	Anxiety	Depression	SC	Impulsive	Vulnerable
Step 1						
ISJT	.08*	.08*	.08*	.08*	.07*	.08*
Neuro facet (NF)	-.02	-.00	-.00	-.02	.00	-.01
$\Delta R^2$	.07 <sup>†</sup>	.06 <sup>†</sup>	.06 <sup>†</sup>	.07 <sup>†</sup>	.06 <sup>†</sup>	.06 <sup>†</sup>
Step 2						
NF $\times$ ISJT	-.01*	-.00	.00	.00	-.01	-.01
$\Delta R^2$	.05*	.00	.00	.00	.02	.02
$R^2$	.12*	.06	.07	.07	.08	.08

*Note.*  $N = 82$ . Job performance DV is standardized. ISJT = interpersonal skills situational judgment test; Neuro = neuroticism; AH = angry hostility; SC = self-consciousness; Impulsive = impulsiveness; Vulnerable = vulnerability to stress. All regression coefficients reported in this table are unstandardized ( $B$ ).

<sup>†</sup> $p < .10$ . \* $p < .05$ .

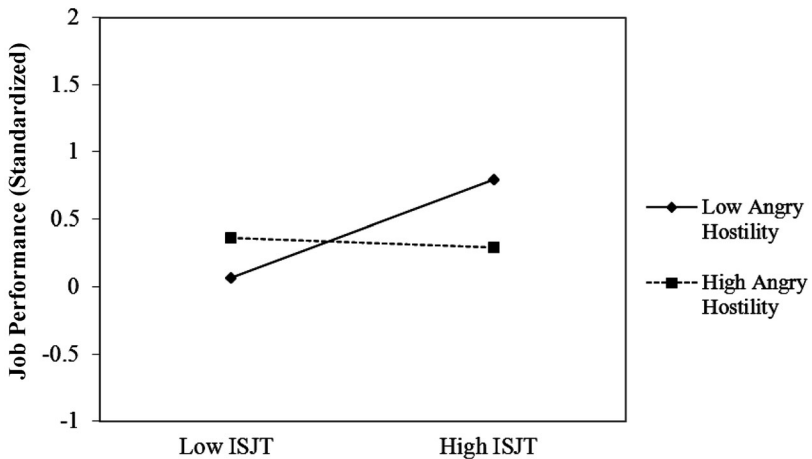


Figure 3: **Interaction of ISJT Score and Angry Hostility Predicting Physician Performance (Study 3).**

*Note.* ISJT = interpersonal skills situational judgment test. High = 1 SD above the mean; Low = 1 SD below the mean.

and negative ( $b = -8.70$ ,  $t = 2.07$ ,  $p = .04$ ). Finally, splitting the sample at the AH median, we also found that the relation between ISJT scores and performance ratings was positive and significant in the low anger group,  $r = .40$ ,  $p < .01$ , and was negative but nonsignificant in the high-anger group,  $r = -.13$ ,  $p = .48$ .

This third study provides a constructive replication (Lykken, 1968) of the hypothesis test, in which the conditions varied from Studies 1 and 2 in a number of ways. First, we utilized a video-based SJT, as opposed to the paper-and-pencil SJTs used in Studies 1 and 2. Second, we used a sample of physicians, a group of employees for whom education level and work tasks differ greatly from those of the police officers in Study 2 and the lower-level healthcare workers in Study 1. Third, we had measures for all of the neuroticism facets in Study 3, as opposed to just a few additional facets in Studies 1 and 2. The AH  $\times$  ISJT interaction was again strong and significant ( $\Delta R^2 = .05$ ), and none of the interactions between ISJT and any of the other neuroticism facets was significant. This is in contrast to the Study 2 finding in which the SC  $\times$  ISJT interaction was significant, but replicates the nonsignificant SC  $\times$  ISJT interaction in Study 1. Taken together, this suggests that this is a rather robust phenomenon that in most contexts is specific to one facet of neuroticism—AH. Moreover, it suggests that it likely exists in a number of different types of jobs of varying complexity and is robust to the more realistic presentation of work scenarios in video-based SJTs (Lievens & Sackett, 2006).

### *General Discussion*

This paper is the first to address personality-based moderators of SJT validity in the context of specific constructs measured. As such, this paper responds to recent calls to address some of the limits of operational SJTs (Ployhart & MacKenzie, 2011; Ployhart & Weekley, 2006) and calls for greater attention to the constructs measured by selection tests (Arthur & Villado, 2008). The results of the three studies support our hypothesis that interpersonally oriented SJTs are less strongly related to job performance among employees who have higher levels of AH. This hypothesis was based on theory and research that suggest that those who are angrier will have more difficulty predicting their future feelings (e.g., Loewenstein, 1996, 2000) and translating their intentions into behavior. This is likely in part because their performance is more variable, and thus less predictable, as these individuals are more reactive in difficult interpersonal situations from a cognitive (Berry et al., 2005), behavioral (Judge et al., 2006), and psychophysiological standpoint (Suls & Wan, 1993). In addition, these findings are consistent with previous research that shows that those higher on AH have an inflated sense of control (Lerner & Keltner, 2001). The results across three studies were consistent with our expectations: A “main effect” of ISJT score, no overall relation between AH and job performance, and an interaction between ISJT and AH, such that the ISJT–performance relationship was considerably stronger among those low on AH.

*Integration With Previous Research Findings*

Our findings suggest that research may benefit from theoretical and empirical work focusing on individual-difference moderators of the relations between many common predictor constructs and job performance. Adopting a “main-effect” perspective in predicting performance is likely an oversimplified view. The results of this study suggest that to be a high performer, one needs more than procedural knowledge (Motowidlo et al., 2006) and good judgment (Brooks & Highhouse, 2006) in interpersonal situations at work. One also needs the ability to translate the procedural knowledge and judgment into behavior on the job.

In the introduction, we discussed work by Chan (2006), who found that SJT scores moderated the relation between proactive personality and job performance. In the present investigations, the finding that AH moderated the relation between ISJT scores and performance is statistically equivalent to the finding that situational judgment moderates the relation between AH and performance. AH was not related to any of the criterion variables in any of the three studies (see Tables 1, 3, and 6). In order to examine the relation in a way that was similar to Chan’s work, we split the Study 1 sample at the ISJT median, creating low ( $n = 98$ ) and high ( $n = 127$ ) ISJT groups. Our findings are similar to those of Chan. Among employees with low ISJT scores, the correlation between AH and performance was positive and approached significance ( $r = .20$ ,  $p = .05$ ), whereas among employees with high SJT scores, the correlation was significant and negative ( $r = -.29$ ,  $p < .01$ ). Thus, when ISJT scores are high, it benefits the organization to choose individuals who also have lower levels of AH.<sup>4</sup>

*Implications and Suggestions for Future Research*

An important theoretical implication of this study concerns the theory of planned behavior (TPB; Ajzen, 1991; Ajzen & Fishbein, 1977). The TPB is relevant to SJTs due to the criticality of the intention–behavior link needed for a strong SJT–performance relationship. Our data suggest support for the proposition of TPB that behavioral control should moderate the relationship between intention and behavior, which is important given that previous empirical tests had been for the most part unresponsive

---

<sup>4</sup>Note that the positive relation between AH and some performance criteria among those with low SJT scores does not translate to a recommendation to select low-SJT applicants who are higher on AH. Rather, the findings suggest that an organization does not gain much by selecting applicants with high SJT scores and higher AH over applicants with low SJT scores and higher AH.

(e.g., Ajzen & Madden, 1986; Doll & Ajzen, 1992). The results of these studies suggest that the moderating influence of behavioral control is likely to be dependent on the objective difficulty of translating intentions to behavior. As we discussed in the introduction, in difficult interpersonal situations at work, it can be quite difficult to exercise self-control and act appropriately, even if one has the appropriate procedural knowledge. This is because it often comes down to “in-the-moment” behavior that is quite different from the behavior examined in typical TPB studies (e.g., voting for a presidential candidate; carrying out planned job search behaviors). Similarly, the one exception to the unsupported tests of the moderating influence of behavioral control mentioned above was a study of weight loss (Schifter & Ajzen, 1985), which is difficult and also dependent upon “in the moment behavior” (i.e., turning down food that does not match one’s diet). When the intention–behavior linkage requires momentary self-control, the ability to control one’s behavior may be a stronger moderator of this relationship.

As noted by one of the reviewers, there does not exist (to our knowledge) an overarching theory that specifically identifies AH as the only source of behavioral variability, unpredictability, and the inability to capitalize on one’s procedural knowledge while simultaneously ruling out the other facets of neuroticism as the potential cause. However, theory and research on incidental anger do suggest that the combination of negative valence and the appraisal of other-control are unique to anger and differ from other negative emotions such as guilt and sadness (Lerner & Keltner, 2001). This leads people who are feeling angry to assign greater blame to others (Keltner et al., 1993) and to distrust other people (Dunn & Schweitzer, 2005). Moreover, we were able to rule out the influence of each of the other facets of neuroticism across the three studies (with the exception of self-consciousness in Study 2). Conceptual and empirical work is still needed, however, to understand the specific role of trait AH in performance variability and behavioral unpredictability. Therefore, this study is best viewed as an empirical building block for future theoretical development on the consequences of trait AH (Locke, 1996).

The findings of this study suggest a number of other potentially interesting avenues for future research. It will be important for future researchers to expand the investigation of the AH facet of neuroticism, both in terms of its role in influencing test scores and as a moderator of the relation between employment assessments and job performance. For example, in these studies, we expected (and found) that AH was only weakly related to ISJT scores ( $r = -.14, .01$ , and  $.04$  in Studies 1, 2, and 3, respectively). This was in large part because the testing methods—responding to multiple-choice questions based on scenarios presented by computer or paper-and-pencil format—made it less likely that trait

AH will cause people to act in an interpersonally inappropriate manner. Thus, in a sense, AH fails to appropriately lower the test taker's interpersonal skills scores when they are measured via SJTs. An important question is whether this is also true of other predictor methods that may be used to assess interpersonal skills, such as interviews (Huffcutt et al., 2001) and assessment centers (Arthur, Day, McNelly, & Edens, 2003). This concern may be minor, given that in interviews and assessment centers candidates are directly demonstrating interpersonal skills with verbal and nonverbal communication. Thus, for these assessments, it is less likely that candidates' AH can be "hidden" from those who judge the test takers' interpersonal skills. Future studies could examine whether AH is related to interview and assessment center measures of interpersonal skills. Our data suggest that when the testing context is an interpersonal one, trait AH is likely to influence others' judgments about the focal person's interpersonal skills. Of course, this should be substantiated empirically, and we urge future researchers to consider how AH directly influences measures of interpersonal skills via the interview and the assessment center.

In terms of practical implications, one could cautiously interpret our findings as suggesting that researchers should perhaps measure AH and use it in combination with ISJT scores when selecting employees. This would mean that individuals would have to be above the cut score for ISJT performance *and* below the cut score on AH to be selected. However, there are a number of issues to consider. First, according to Society for Industrial and Organizational Psychology's (SIOP's) *Principles for the Validation and Use of Employee Selection Procedures* (2003), "An extremely large dataset or replication is required to give full credence to unusual findings [such as] . . . moderator effects" (p. 21). We have replicated the AH  $\times$  ISJT finding in three different studies, each with different samples and settings, but we recommend that practitioners conduct local validation studies to replicate this interaction. Second, there is an important issue regarding the use of a moderator variable that influences the relationship between the ISJT and job performance but does not itself correlate with performance (AH was not significantly correlated with job performance in any of the these three studies) and has not been established through professional job analysis as an important trait for job performance. For these reasons, we cannot advocate for the use of AH as a predictor at this time. However, we would recommend that entities responsible for producing guidelines on selection procedures (e.g., SIOP members charged with the next revision of the *Principles*) think through the possibility that such predictors may be appropriate to include in selection batteries, in light of how drastically they may influence the validity of inferences drawn from scores on other predictor measures.

A third and final issue to consider from a practical standpoint is that, in each of the studies we conducted, the personality test administration was separate from the selection testing, so participants were less motivated to “fake good” on the AH scale than they would be when applying for a job. It is possible that the results we observed would not have been revealed in a situation in which applicants are motivated to manage impressions on personality scales. Therefore, one important direction for future research is to determine whether these personality findings are also observed in applicant samples where personality is measured as part of the selection battery. Future research could also employ different ways of measuring AH that might be less subject to response distortion, such as hostile attribution bias (Matthews & Norris, 2002) and conditional reasoning measures (e.g., LeBreton, Barksdale, Robin, & James, 2007).

### *Limitations and Conclusions*

There are a number of limitations of this investigation that should be addressed. The first is that these studies utilized supervisor ratings of performance. It is possible that angry and hostile behavior might have negatively influenced the supervisory ratings. However, the results do not suggest that this was a problem in any of the studies, as AH was unrelated to the performance criteria. A second potential concern to note is the long delay between ISJT measurement in Studies 2 (two years) and 3 (nine years). Particularly in Study 3, the criterion measure is temporally distal. It should be noted, however, that in both studies the relations between ISJT scores and performance were significant among those low in AH (.51 in Study 2, .40 in Study 3), which suggests the appropriateness of the criterion.

A third limitation concerns the construct validity of ISJTs in this study. For example, the low internal consistencies for the ISJT measures are in line with prior research on SJTs (e.g., Whetzel & McDaniel, 2009) because SJT items are typically construct heterogeneous at the item level. SJT items and responses are typically derived from critical incidents that reflect important performance domains, which may draw on numerous knowledge, skills, and abilities (e.g., McDaniel et al., 2001; Schmitt & Chan, 2005). Although this leads SJTs to be somewhat “impure” measures of single constructs, they are often saturated with predominant higher-order constructs (Christian et al., 2010; Roth, Bobko, McFarland, & Buster, 2008). Despite evidence that the ISJTs in our study were saturated with interpersonal skills (SME ratings in Study 1, convergent validity correlations in Study 2), we caution against the overinterpretation of our results, as each of the ISJTs is likely to measure other job-related constructs, albeit to a lesser degree.



A fourth limitation is the small sample size on which the key findings of Study 2 are based ( $N = 54$ ). Caution should be exercised when making conclusions on the basis of results from samples that are of this size. A post hoc analysis showed that we only had power = .45 to detect the observed AH  $\times$  ISJT interaction, given the sample size and the size of the effect (by comparison, the power to observe the hypothesized interactions was .85 in Study 1 and .62 in Study 3). Importantly, the AH  $\times$  ISJT interaction observed in Study 2 was replicated in each of the other studies. However, the SC  $\times$  ISJT interaction was replicated in neither Study 1 nor Study 3, and thus, we are less confident in making any strong conclusions about this interaction.

A fifth limitation,<sup>5</sup> or at least a potential limitation, is the low means and the small standard deviations for AH in all three studies. Thus, people who were classified in the study as high AH (e.g., when we calculated simple slopes or presented graphs of the interaction) would perhaps be more appropriately categorized as moderate AH. This suggests one of two possibilities: Either participants engaged in some form of response distortion and they are truly higher on AH than their self-rated scores reflect, or we managed to sample only low- to moderate-AH employees and applicants (i.e., it is possible that “true” high-AH employees exist in the population but did not make it into our sample). First, consider the possibility that participants in our study were distorting their responses. This would either have no effect on the results (if each participant was distorting his or her responses at the same level and in the same direction) or would serve to make the observed results look weaker than they actually are (if participants were distorting their responses at different levels, leading to different rank orders for true scores and observed scores). Second, consider what would happen if there were no distortion but we somehow systematically failed to sample people who are truly high AH: What would the ISJT–performance relationship be for those who are truly high on AH? One possibility is that there is a complex, quadratic moderated relationship: The ISJT–performance relationship is strong for people low on AH; weak (near zero) for people moderate on AH; and strong again for people on AH. Although this is possible, our theory suggests that the more likely result is that for people who are truly “high AH,” the relationship between ISJT and performance is also likely to be weak. That is, the variability and unpredictability of people high on AH is likely to be stronger than for people who are moderate on AH. Again, this suggests that our results (and hence, our conclusions) are conservative, and the problem with ISJTs is larger than we have described.

---

<sup>5</sup>We are grateful to an anonymous reviewer for pointing this out to us.

In conclusion, these studies identified AH as an important influence on the validity of ISJTs. We are hopeful that these findings will help to stimulate future research in a variety of areas, including SJTs as measures of specific types of procedural knowledge, the relevance of the theory of planned behavior to SJT–performance relationships, and individual-difference variables that moderate the relationship between predictor constructs and job performance. From a practice perspective, an important next step will be to identify applicants during the selection process, for whom inferences from ISJT scores may not be valid.

## REFERENCES

- Aiken LS, West SG. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.
- Ajzen I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211.
- Ajzen I, Fishbein M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84, 888–918.
- Ajzen I, Madden TJ. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, 22, 453–474.
- Arthur W, Jr., Day EA, McNelly TL, Edens PS. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *PERSONNEL PSYCHOLOGY*, 56, 125–154.
- Arthur W, Jr., Villado AJ. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435–442.
- Ausubel LM. (1991). The failure of competition in the credit card market. *American Economic Review*, 81, 50–81.
- Bazerman MH, Tenbrunsel AH, Wade-Benzoni KA. (1998). Negotiating with yourself and losing: Making decisions with competing internal preferences. *Academy of Management Review*, 23, 225–241.
- Berry JW, Worthington EL, Jr., O'Connor LE, Parrott L, III, Wade NG. (2005). Forgiveness, vengeful rumination, and affective traits. *Journal of Personality*, 73, 183–225.
- Boatman J, Erker S. (2012). *Global selection forecast 2012*. Bridgeville, PA: Development Dimensions International.
- Brooks ME, Highhouse S. (2006). Can good judgment be measured? In Weekley JA, Ployhart RE (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 39–56). Mahwah, NJ: Erlbaum.
- Carpenter TD, Wisecarver MM. (2004). *Identifying and validating a model of interpersonal performance dimensions*. Technical Report 1144, United States Army Research Institute for the Behavioral and Social Sciences.
- Caspi A, Elder GH, Bem DJ. (1987). Moving against the world: Life-course patterns of explosive children. *Developmental Psychology*, 23, 308–313.
- Catano VM, Brochu A, Lamerson CD. (2012). Assessing the reliability of situational judgment tests used in high-stakes situations. *International Journal of Selection and Assessment*, 20, 333–346.

- Chan D. (2006). Interactive effects of situational judgment effectiveness and proactive personality on work perceptions and work outcomes. *Journal of Applied Psychology*, 91, 475–481.
- Christian MS, Edwards BD, Bradley JC. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *PERSONNEL PSYCHOLOGY*, 63, 83–117.
- Clause CS, Mullins ME, Nee MT, Pulakos E, Schmitt N. (1998). Parallel test form development: A procedure for alternative predictors and an example. *PERSONNEL PSYCHOLOGY*, 51, 193–208.
- Costa PT, McCrae RR. (1992). *NEO-PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa PT, McCrae RR. (1995). Domains and facets: Hierarchical personality using the revised NEO Personality Inventory. *Journal of Personality Assessment*, 64, 21–50.
- Davis MC, Matthews KA, McGrath CE. (2000). Hostile attitudes predict elevated vascular resistance during interpersonal stress in men and women. *Psychosomatic Medicine*, 62, 17–25.
- Doll J, Ajzen I. (1992). Accessibility and stability of predictors in the theory of planned behavior. *Journal of Personality and Social Psychology*, 63, 754–765.
- Dunn JR, Schweitzer ME. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, 88, 736–748.
- Ferris GR, Witt LA, Hochwarter WA. (2001). Interaction of social skill and general mental ability on job performance and salary. *Journal of Applied Psychology*, 86, 1075–1082.
- Fredrickson BL, Maynard KE, Helms MJ, Haney TL, Siegler IC, Barefoot JC. (2000). Hostility predicts magnitude and duration of blood pressure response to anger. *Journal of Behavioral Medicine*, 23, 229–243.
- Gardner H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.
- Gold R. (1993). On the need to mind the gap: On-line versus off-line cognitions underlying sexual risk-taking. In Terry D, Gallois C, McCamish M (Eds.), *The theory of reasoned action: Its application to AIDS preventive behavior* (pp. 227–252). New York, NY: Pergamon.
- Goldberg LR. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In Mervielde I, Deary I, De Fruyt F, Ostendorf F (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Hoekstra HA, Ormel J, De Fruyt F. (1996). *NEO persoonlijkheidsvragenlijsten NEO PI-R en NEO-FFI*. Lisse, The Netherlands: Swets & Zeitlinger.
- Hu L, Bentler PM. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Huffcutt AI, Conway JM, Roth PL, Stone NJ. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86, 897–913.
- Judge TA, Scott BA, Ilies R. (2006). Hostility, job attitudes, and workplace deviance: Test of a multilevel model. *Journal of Applied Psychology*, 91, 126–138.
- Keltner D, Ellsworth PC, Edwards K. (1993). Beyond simple pessimism: Effects of sadness and anger on social perception. *Journal of Personality and Social Psychology*, 64, 740–752.
- Klein C, DeRouin RE, Salas E. (2006). Uncovering workplace interpersonal skills: A review, framework, and research agenda. In Hodgkinson GP, Ford JK (Eds.),

- International review of industrial and organizational psychology* (Vol. 21, pp. 80–126). New York, NY: Wiley & Sons, Ltd.
- Lazarus RS. (1991). *Emotion and adaptation*. New York, NY: Oxford University Press.
- LeBreton JM, Barksdale CD, Robin J, James LR. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology*, 92, 1–16.
- Lerner JS, Keltner D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, 81, 146–159.
- Lerner JS, Tiedens LZ. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of Behavioral Decision Making*, 19, 115–137.
- Lievens F, Buyse T, Sackett PR. (2005a). Retest effects in operational selection settings: Development and test of a framework. *PERSONNEL PSYCHOLOGY*, 58, 981–1007.
- Lievens F, Buyse T, Sackett PR. (2005b). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90, 442–452.
- Lievens F, Coetsier P, De Fruyt F, De Maeseneer J. (2002). Medical students' personality characteristics and academic performance: A five-factor model perspective. *Medical Education*, 36, 1050–1056.
- Lievens F, Ones DS, Dilchert S. (2009). Personality scale validities increase throughout medical school. *Journal of Applied Psychology*, 94, 1514–1535.
- Lievens F, Peeters H, Schollaert E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37, 426–441.
- Lievens F, Sackett PR. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91, 1181–1188.
- Lievens F, Sackett PR. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, 97, 460–468.
- Locke EA. (1996). Using programmatic research to build grounded theory. In Frost PJ, Taylor MS (Eds.), *Rhythms of academic life* (pp. 99–105). Thousand Oaks, CA: Sage.
- Loewenstein G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65, 75–98.
- Loewenstein G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review*, 90, 426–432.
- Loewenstein G, Schkade D. (1999). Wouldn't it be nice? Predicting future feelings. In Kahneman D, Diener E, Schwarz N (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 85–105). New York, NY: Sage.
- Lykken DT. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70, 151–159.
- Lynch BS, Bonnie RJ. (1994). Toward a youth-centered prevention policy. In Lynch BS, Bonnie RJ (Eds.), *Growing up tobacco free: Preventing nicotine addiction in children and youths* (pp. 3–25). Washington, DC: National Academy Press.
- Mathieu JE, Gilson LL, Ruddy TM. (2006). Empowerment and team effectiveness: An empirical test of an integrated model. *Journal of Applied Psychology*, 91, 97–108.
- Matthews BA, Norris FH. (2002). When is believing "seeing?" Hostile attribution bias as a function of self-reported aggression. *Journal of Applied Social Psychology*, 32, 1–32.

- McDaniel MA, Hartman NS, Whetzel DL, Grubb WL, III. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *PERSONNEL PSYCHOLOGY*, 60, 63–91.
- McDaniel MA, Morgeson FP, Finnegan EB, Campion MA, Braverman EP. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730–740.
- Milgram S. (1965). *Obedience to authority*. New York, NY: Harper and Row.
- Moorman RH, Blakely GL. (1995). Individualism-collectivism as an individual-difference predictor of organizational citizenship behavior. *Journal of Organizational Behavior*, 16, 127–142.
- Motowidlo SJ, Dunnette MD, Carter GW. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 640–647.
- Motowidlo SJ, Hooper AC, Jackson HL. (2006). A theoretical basis for situational judgment tests. In Weekley JA, Ployhart RE (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 57–81). Mahwah, NJ: Erlbaum.
- Mumford MD, Peterson NG. (1999). The O\*NET content model: Structural considerations in designing jobs. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O\*NET* (pp. 21–30). Washington, DC: American Psychological Association.
- Murphy BC, Eisenberg N. (1997). Young children's emotionality, regulation, and social functioning and their responses when they are targets of a peer's anger. *Social Development*, 6, 18–36.
- Ployhart RE, MacKenzie WI, Jr. (2011). Situational judgment tests: A critical review and agenda for the future. In Zedeck S (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 2, pp. 237–252). Washington, DC: American Psychological Association.
- Ployhart RE, Weekley JA. (2006). Situational judgment: Some suggestions for future science and practice. In Weekley JA, Ployhart RE (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 345–350). Mahwah, NJ: Erlbaum.
- Plutchik R. (2003). *Emotions and life*. Washington, DC: American Psychological Association.
- Roth P, Bobko P, McFarland L, Buster M. (2008). Work sample tests in personnel selection: A meta-analysis of black-white differences in overall and exercise scores. *PERSONNEL PSYCHOLOGY*, 61, 637–661.
- Schifter DE, Ajzen I. (1985). Intention, perceived control, and weight loss: An application of the theory of planned behavior. *Journal of Personality and Social Psychology*, 49, 843–851.
- Schmidt S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13, 90–98.
- Schmitt N. (2008). The interaction of neuroticism and gender and its impact on self-efficacy and performance. *Human Performance*, 21, 49–61.
- Schmitt N, Chan D. (2005). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage.
- Society for Industrial and Organizational Psychology (SIOP). (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.
- Sternberg RJ, Conway BE, Ketron JL, Bernstein M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology*, 41, 37–55.
- Suls J, Wan CK. (1993). The relationship between trait hostility and cardiovascular reactivity: A meta-analysis. *Psychophysiology*, 30, 615–626.

- Tellegen A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In Tuma AH, Maser JD (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Erlbaum.
- Tiedens LZ, Linton S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social Psychology*, 81, 973–988.
- Tsang EW, Kwan KM. (1999). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review*, 24, 759–780.
- Van Dyne L, LePine JA. (1998). Helping and voice extra-role behavior: Evidence of construct and predictive validity. *Academy of Management Journal*, 41, 108–119.
- Weekley JA, Ployhart RE. (Eds.) (2006). *Situational judgment tests: Theory, measurement and application*. Mahwah, NJ: Erlbaum.
- Whetzel DL, McDaniel MA. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19, 188–202.
- Whetzel DL, McDaniel MA, Nguyen NT. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21, 291–309.
- Williams LJ, Anderson SE. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, 17, 601–617.

## APPENDIX

### Example ISJT Items in Studies 1, 2, and 3

#### *Study 1 Example Item*

An eight-year-old male patient is engaged in an irritating behavior. The behavior is harmless, but other patients are paying attention to it. He ignores you when you ask him politely to stop the behavior. How would you respond at this point?

- (a) Ignore the behavior and begin a new group activity.
- (b) Tell the patient there will be consequences if he continues his behavior.
- (c) Tell the patient, “I am in charge here, and you need to stop.” If the behavior continues, physically remove him from the room.

#### *Study 2 Example Item*

You are assigned as one of two lieutenants in a patrol division. You are responsible for supervising half of the sergeants in the division. You have a peer lieutenant, senior to you, who supervises the other half of the sergeants. You are approached by one of your sergeants who complains to you that your peer is always monitoring his performance and checking up on him. He informs you that he works for you and he will not tolerate this behavior from your peer lieutenant. You are aware that your peer is

very attentive to call load management by sergeants. The sergeant asks that you intervene.

What would you most likely do?

- (a) Tell the sergeant that the peer lieutenant has as much authority to supervise him as you do.
- (b) Meet with your peer lieutenant to discuss the matter.
- (c) Advise your peer lieutenant to come directly to you with issues they may have with your sergeants.
- (d) Advise your sergeant to ignore the peer lieutenant as the sergeant is responsible to you only and you will take care of any issues.
- (e) Inform the captain of the situation and ask for direction.

*Study 3 Example Item*

Physician: I am going to prescribe some medication that should substantially improve the symptoms in the next days or so.

Patient: Hopefully. Yes.

Physician: I will prescribe antibiotics.

Patient: Antibiotics? I'm actually opposed to taking antibiotics. People say they are no good. I'd rather not take them.

What is the best way for you (as a physician) to react to this patient's refusal to take the prescribed medication?

- (a) Tell him that, in his own interest, it is important that he take the antibiotics.
- (b) Clarify in a friendly way that such an attitude will not solve his problems.
- (c) Explain that all scientific experts agree that antibiotics are needed here.
- (d) Emphasize that his problems will not go away without antibiotics.